DOCUMENT RESUME

ED 227 131 TM 830 119

AUTHOR Hills, John R.; King, F. J.
TITLE Construct Validity of the Florida Functional Literacy Test.
INSTITUTION Florida State Univ., Tallahassee. Coll. of Education.
SPONS AGENCY Florida State Dept. of Education, Tallahassee. Student Assessment Section.
PUB DATE 25 Aug 78
CONTRACT SDE-780-090
NOTE 50p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adults; *Communication Skills; *Functional Literacy; Graduation Requirements; High School Students; *Mathematics Skills; *Minimum Competency Testing; Reading Comprehension; Scores; Secondary Education; Socioeconomic Status; State Programs; Testing Programs; *Test Validity
IDENTIFIERS *Florida Functional Literacy Test

ABSTRACT

The 1977 edition of the Florida Functional Literacy Test is a 117-item multiple-choice measure of 13 mathematics and 11 communication skills required for minimal acceptable performance for students to graduate from state high schools. A student must pass at least half of the skills in mathematics and communications and get 70 percent of the items for each skill correct. To investigate the test's construct validity, the literacy test performances between groups of adults at three education and income levels were compared with performace on an independent multiple-choice cloze test of communications skills in reading. The groups also completed demographic descriptions. A substantial correlation between the communications total score on the literacy test and the independent measure of reading ability indicated strong support for the test's validity. The proportion passing the mathematics standard and communications standard increased markedly from the low- to high-level functional groups. Mean scores behaved similarly, indicating strong statistical support for validity. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED227131

CONSTRUCT VALIDITY OF THE FLORIDA FUNCTIONAL LITERACY TEST

BY

John R. Hills

AND

F. J. King

Assisted by Joan Gallini and David Logsdon

Educational Research, Development, and Foundations

College of Education, Florida State University

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Under Contract to the Assessment Section,
Bureau of Program Support Services,
Florida Department of Education
SDE #780-090

August 25, 1978

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. Ferrell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Copyright
State of Florida
Department of State
1978

TM 830 119

NEED FOR VALIDATION

The Florida Functional Literacy Test (1977 edition) is a 117 item multiple-choice test using a separate, machine-scorable answer sheet, which measures mathematics and communications skills appropriate for minimal acceptable performance of students graduating from Florida high schools. There are 13 separate mathematics skills and 11 separate communications skills. Each of those skills is measured by several items, with a specified score set for passing each of them. A student must pass at least half of the communications skills to pass the communications standard, and he must get 70% of the communications items correct. He must perform similarly well in mathematics. He must have mastered both standards to be awarded a diploma. The requirement for this testing program was established by the Florida Legislature and was implemented in this form by the Assessment Section of the State Department of Education.

Content validity is a very important characteristic of an achievement test. The content validity of any achievement test is a matter of how well the items reflect the content which the test is supposed to measure. It is usually established by careful specification of the test objectives, by careful item writing and editing, and by careful review of the finished product to be sure that it still reflects

accurately what it was intended to reflect.

A properly-trained teacher using sound testing procedures specifies her objectives and designs items to fit them, thus providing content validity. However, many teachers' tests used for giving grades and deciding pass and fail are not even evaluated for (or designed to have) content validity. Too many teachers have not been trained adequately in this phase of their profession. It is also true that many licensing board exams are not checked for content validity, even though careers and livelihoods depend on their scores. But these are conditions to be deplored rather than emulated.

The Florida Functional Literacy Test has been carefully designed to have content validity. The objectives for the test were painstakingly developed, and items were prepared to reflect those objectives. Thus, the test has been more carefully developed than many educational tests and most certification tests.

Given the context in which the Literacy Test was conceived--i.e., the application of certain basic skills to practical problems as a final examination for high school students--adequate content validity permits interpretation about the degree to which students have mastered the intended requirements. However, additional information about the test should be continually sought as

a means of maintaining quality and providing reassurance that the test scores behave as one would expect.

This falls into the realms of correlational and construct validity. Here such questions are asked as whether the scores from the test under question tend to agree with scores from other tests that should be related to it and fail to agree with or are unrelated to measurements of quite logically distinct characteristics. The scores should be related to characteristics of examinees other than test scores, such as ages among school students, levels of education, and similar variables. Presumably, one should be able to change certain characteristics of the examinees, such as teaching them specific content, and such changes should be reflected in the test scores in anticipated ways. One should even be able to make predictions of future events or scores on the basis of the scores of the test under question and find those predictions verified to a large extent if the scores are behaving as they should. These are the issues of construct validity. When a series of observations of these various kinds has been assembled for a test, one feels that he has a network of supportive relationships that add meaning to and understanding of the test scores--relationships which reassure that community that the scores can be trusted.

Obviously, establishing such a network is no easy or

short-term thing. The more important the test, the more extensive the validation should be, the longer it will take, and the more it will cost. Unfortunately, validation is never concluded; there are always more questions to ask and more relationships to explore.

Before it is finished, the network of relationships should include such things as the relationship between these test scores and previous school grades, current school grades, and future school and college grades. These are some of the variables that should reasonably be related to literacy test scores. The post high-school activities of students should be related to their Literacy Test scores. Presumably those who soon find themselves in substantial positions in the community with large incomes and important influence should be the ones with higher scores, and those with lower scores should be less influential. The Literacy Test scores should, one would think, be related to the Basic Skills Test scores that are also in use in Florida. The rationale is that the literacy test measures the application of those skills to real-life problems, and those who have mastered the skills to the greatest degree should be able to apply them with the greatest ease.

Nature of this Validation

Obviously, it would demand great resources to push forward all aspects of validation at once. In fact, as one works on one type of validation, other possibilities occur. There exists no catalog of all possible kinds of data that might be used to support or bring into question or reinterpretation the meaning of test scores. But one must start somewhere, and it seemed that a very important initial question for a test with the intent of the Functional Literacy Test was that of how well the scores correspond with the level at which different people function in society.

It made no sense to act as though people who scored low on such a test necessarily would not be able to function--would not be able to earn a living, vote, operate an automobile, etc. All of us know of people who earn satisfactory livings who cannot be left a written message because they cannot read it. Some of us know mechanics, tile setters, carpenters, and others who are quite successful through use of methods that do not require the usual kinds of mathematics and methods that were learned by example and word of mouth rather than through reading. Delivery men function without reading skills by asking people to read names, addresses, and maps for them. However, it seems that if one were to test a group of people who were

employed in low level occupations, earned relatively low incomes, and had a relatively low level of education, and if he were also to test people in good-paying jobs, in occupations of relatively high level, and who had graduated from high school, the scores on a sound functional literacy test should be noticeably different, on the average, in those two groups. So we set out to try to locate such groups, test them, and compare their performances on the Florida Functional Literacy Test.

Functional literacy as the concept is being used in this context has been defined as follows: Functional literacy is the satisfactory application of basic skills in reading, writing, and arithmetic to problems and tasks of a practical nature as encountered in everyday life.

Now a simple and appealing idea is that a person's teacher has a great deal of contact with that pupil in matters related to literacy as defined. If the data could be easily obtained, it would be interesting to see whether teachers could judge their individual pupils' levels of literacy and whether those judgments corresponded with passing and failing the Literacy Test.

Finally, we had available a well-developed, carefully-studied, reading comprehension test designed for about the level of communications skill that represented passing the Literacy Test and whose content was clearly the application

of basic skills to real-life problems. We decided to ask our subjects to take that reading comprehension test as well as the Literacy Test to see whether the scores corresponded as they should if the Literacy Test is measuring what it was designed to measure. An additional attractive feature of the reading comprehension test used here is that due to the nature of its development it is possible to ascertain through it the kind of books, magazines, and other material that a person who just passes the Communications Standard on the Literacy Test reads with minimal comprehension. That knowledge should help everyone understand just what level of reading skill is being demanded when the Literacy Test is used as a graduation requirement.

Thus in this validation study it was intended that three criteria or comparisons be used. The Literacy Test performances were to be compared between groups of people who apparently function at different levels in society. They were to be compared with teachers' evaluations of literacy. And they were to be compared with performance on another measure of communications skills in reading. If the scores behaved as one should expect them to in these three settings, a reasonable and reassuring beginning of a network of construct validity would be established.

The Three Criteria

The easiest criterion to discuss is the teacher rating. Each student was presented with a sheet which asked him for certain relevant demographic data. (The sheet appears in Appendix A.) The examinee filled out this sheet. When he had finished, he was asked to take the reading comprehension test, which was attached to the demographic information sheet by a staple. At the conclusion of that, the Literacy Test booklets and answer sheets were distributed. Each Literacy Test had on it a serial number. Examinees were asked to copy that serial number onto their demographic information blank and onto their answer sheet. Thus all three elements shared for each examinee a unique identification number. Then the demographic information blanks and reading comprehension tests were collected, and the teacher was asked to judge whether each student was functionally literate in mathematics and in communication, marking his judgments in spaces provided in the upper right hand corner of the demographic information sheet. The teacher was given a copy of the State's definition of functional literacy, presented above, to assist in making these judgments.

During the initial planning of the study it seemed that it would be simple to collect this information for each subject. It turned out that when arrangements were made for testing in the various centers around the State, the local coordinators

often combined several teachers' students into one large room for testing. In some cases, the teachers were dismissed, since they were not needed during the testing. Thus, for many students the teachers were not available to make the ratings. Further, it turned out that at the upper level of functioning, a large number of the examinees were employed adults, not participants in school classes, and thus not under the supervision of a teacher who could make such judgments. As a result of these complications, this criterion was abandoned. It simply was not feasible under the conditions of this study.

The second criterion was levels of functioning. The work on the Adult Proficiency Level project of Northcutt (Northcutt, et al., 1975) influenced us to take into consideration three aspects in determining level of functioning. First, we considered level of education, asking each examinee to indicate on the demographic information blank the number of years of schooling he had completed. Second, we asked the person to tell us what kind of work he did. Some were able to do that reasonably well, but a large number of our final group (57) called themselves housewives, and another large number (186) were unemployed. We used decile ratings from Duncan's system (Robinson, et al., 1969) to classify the jobs of the 158 employed people, with two raters making the ratings independently and differences

resolved by a third rater. The third was the annual income of the home.

The third criterion was performance on a reading comprehension test. The reading comprehension test used in this study was developed by one of the authors in conjunction with another research program aimed at the educable mentally retarded. The test developed there has a number of unique features. Only the highest level parts of it were used in this study for a group of examinees with a higher level of performance.

The test format is multiple-choice cloze. In a multiple-choice cloze test, a student is given a carefully-selected paragraph to read. In that paragraph, periodically the student is given four words from which he is to choose the one that makes the best sense in the paragraph. The words from which he must choose are words which fit the paragraph grammatically but don't make sense in the context, except for the correct one. For example, in the paragraph below the underlined word makes sense, but the other words could fit, are the same part of speech, etc., but just don't make good sense in light of the rest of the paragraph.

11.

A plan for spending [money / choices time / loans] is called a budget. Sometimes a [budget / purchase washer / carton] is easier to make [whether / than for / rather] it is to keep. Many [orders / budgets people / banks] have problems because they [call / build hurry / buy] more than they can afford.

Of course, one could use a "pure cloze" test in which each subject had to write the correct word, rather than choose the correct word, but it was decided to use the multiple-choice format as one which was more economical, was sufficiently effective for our purposes, and one which, in this case, had an extensive background of evidence concerning its effectiveness and the meaning of its scores.

The cloze test used here included four kinds of content, employability skills, general health, personal money management, and food purchasing and preparation which had been chosen from among seventeen different topics judged of importance for even the educable mentally retarded to be able to read. A separate cloze test was used for each content area.

For each of those content areas the cloze test was

constructed by first collecting from the environment a wide variety of materials available to the general public on that topic. For example, pamphlets distributed on health topics in doctors' offices were gathered, articles from the local newspaper, labels from food containers, etc. A passage of about 200 words in length from each document was analyzed for reading difficulty following the standard procedures as synthesized by King (1974). These were then assembled into forms with strata determined by difficulty level in such a manner that students who passed any particular difficulty level had a very high probability level of also passing all lower levels of difficulty. Once students fail at any level of difficulty, they are very unlikely to pass any higher level of difficulty. This provides essentially a Guttman scale (1950), a very attractive measurement characteristic. Further, since the passages had been stratified by reading difficulty level, one could say for any student who passed at say the fifth level, he could read with comprehension other materials at that difficulty level. Previous analyses of more widely-known material, such as widely read books -- The Last of the Mohicans, The Call of the Wild and Treasure Island --

let one describe the level on the reading comprehension test in terms of these kinds of widely-known literature. This should be very helpful in enabling the layman to comprehend the meaning of the level of performance represented by the passing score on the Functional Literacy Communications Standard.

Data Collection and Sample

The objective of the data collection phase of the project was to obtain substantial numbers of cases at three levels of societal functioning. One level was to have little discretionary income, be unemployed or in a job at a low socio-economic level according to Duncan's index (Robinson, et al., 1969), and have a low level of formal education. The second level was to be approximately at the middle on each of these indices. The third level was to be relatively high on each of these. By relatively, we mean that we expect Functional Literacy scores to differ noticeably between the three levels. By substantial numbers of cases, we mean that if the numbers of cases we gather do not show noticeable differences between pairs of these levels, then whatever differences there might be are educationally of little importance. Approximately thirty cases at each level should be sufficient. We expect that a Functional Literacy Test whose scores are markedly different for people at the low, at the middle, and at the high levels of our indices,

has scores which are behaving as such test scores should behave in this respect. It is not necessary for this purpose that the people included in the study be representative of people in general, or of people in these particular strata. It would be nice if they were representative. Then one could describe the scores that are to be expected from people in each group with a certain margin of error. But that was not our objective, and to be able to do that would be far more expensive. So we merely set out to find people who could be categorized reasonably well into our levels, and we tried to obtain subjects from widely-dispersed parts of the State and from urban and rural systems. We wanted to avoid the possibility of undue local influence.

We were not able to achieve our goal entirely. It turned out to be very difficult to find people at the upper level who would give up several hours of time as a group to participate in this study. Finally we resorted to a special group, people employed in the State Department of Education, to provide cases at our high level of functioning. We did test people in varied locations, as shown in the chart below:

Description of Persons Tested

Region	Level of Program	Number	Percent
Northwest	Adult Basic & Secondary	138	28
North Central	Adult Basic	13	3
	" Secondary	13	3
	Employed Adults	48	10
Northeast	Adult Basic	46	9
West Central	Adult Basic	56	11
Southeast	Adult Basic & Secondary	104	21
Southeast	Community College	71	14
No Record for Region		4	1
TOTAL		493	100

The result was a matrix of subjects in various categories from which we chose selected cells for our three levels. We found the job classification to be of little use (see page 36), so that index was not used to categorize subjects. The matrix of subjects by levels in the remaining two indices appears in Table 1. Subjects in cells marked A were used in our low group, in B our middle group, and in C our high group.

A separate analysis was made of the scores on the reading comprehension cloze test and the scores on the Communications Standard of the Functional Literacy Test. We

Table 1

Criterion Groups
By Education By Income

Income \ Education	0 - 8 yrs	9 - 11 yrs	12 & more	TOTALS
Less than $5000	A 40	32	7	79
$5000 - $14,000	65	B 75	42	182
More than $14,000	28	57	C 70	155
Examinee failed to report education or income.			32	32
				448

wanted to establish that there was a substantial degree of correlation between those two tests. Since the cloze test is a well-recognized and highly respected measure of reading skill, and since this particular cloze test produces scores of a high degree of validity in the form of a Guttman scale, the degree to which the Functional Literacy scores correlate with the cloze test scores should indicate how well the Literacy Test scores measure reading skill. Subjects in cell B were used in that analysis. These were chosen because they were subjects who could be expected to read with skill reasonably well reflected in the score range of the cloze test.

Some modifications of the standard testing procedure for the Functional Literacy Test were made for this study. When the test was administered in the state-wide testing in schools, unlimited time was allowed for students to take the test. It is rumored that a few students even took seven or eight hours. That was manifestly impossible for adults. Most often we obtained our adults from school settings in evening classes in which they had previously scheduled themselves to be picked up or to catch a bus after about a two-hour period. Similar student scheduling and transportation problems occurred in the afternoon testing of adults. We did not expect many to desire to work on these tests more than two or two and one-half hours, at any rate. We surmised that

the low-scoring students would get bored and stop, and the high-scorers would not take more than an hour and a half to two hours.

In order to explore some of these ideas we tested two small pilot groups of pupils in an adult basic and an adult secondary education class in North Central Florida. There we noted the scheduling problem, and we also noted that these adult basic education students had a considerable degree of difficulty with the Functional Literacy Test. One of them spent his entire time on the first section, never reaching the second section on communications skills. We also discovered that the cloze test was very difficult for the adult-basic-education group, few of them getting beyond the first level. (The person who never got to the communications section was not used in further data analyses.)

In the interests of making the testing as palatable as possible and as informative to the study as possible, two changes in procedure were introduced after the pilot testing. First, we decided that for adult-basic-education classes, we would administer only the first section of the cloze test. Second, we decided that for the Functional Literacy Test we would modify the timing. We decided to give the students 45 minutes for the mathematics section, followed by 45 minutes for the communications section, then followed by unlimited time to work on any questions they wanted. This would assure us that at least some work would

be done on the communications section by all students who followed the instructions. We anticipated that the low-scoring students would have revealed all they could do on the mathematics section after 45 minutes, and the same would be true on the communications section in a similar amount of time. However, the privilege of returning after an hour and a half to work on any part they wanted to would preserve the essential characteristics of the standard administration. We deviated from this procedure of timing for the one group of employed adults who were expected to be high scorers due to their jobs and salaries. We expected them all to finish in the two hour block of time we were permitted to use the testing room, and most of them did so with no trouble.

It is significant to note that changes such as these must be made for this kind of validation study of the Functional Literacy Test--the adult world to which we wish to generalize at least tentatively is a different world from the school world, and if the study is to be done at all, it must accommodate to reality. However, the accommodations we made should have very little influence on the main goal of the study, determining whether people functioning at a lower level in society on the average are less likely to pass the Literacy Test standards and are likely to earn lower scores.

We made another modification in procedure after testing our second adult-basic-education group. Those people performed much better on the cloze test than had the pilot group. As a result, we reversed our earlier decision and from then on administered the complete cloze test to all groups. (The group who took the shortened cloze test were eliminated from analyses involving that test.) This sequence of modifications was felt to have little significant effect on the research program. The function of the cloze test was two-fold. One goal was to ascertain the degree of correlation between the Functional Literacy Test scores and the cloze test scores which would tell us to what degree communications skills as measured on the Literacy Test corresponded to reading skills as measured by another test. The other was to obtain an estimate of the level of reading that minimally competent students could comprehend. We planned to do those analyses on the middle-level group, so the performances of the adult basic education group subjects would not be heavily involved (if they were included at all) in those phases of the study.

Another modification that was used in this study concerned the answer sheets. The contractor who prepared the Functional Literacy Test for state-wide administration arranged for the alternative responses to items to be lettered, rather than numbered, and it used the letters

ABCD and FGHI for response options to alternate items. The standard answer sheets for the test were not scorable by the locally-available optical scanner, and it was inconvenient to send the answer sheets from this experiment to a subcontractor for scoring. Therefore we decided to use answer sheets scorable by the local scanner. These had the response options all labeled 1, 2, 3, 4, or 5. To make the transition as easy for the students as possible, in each administration the administrators wrote on the blackboard in at least two places in very large letters and numbers the translation from ABCD FGHI to 1, 2, 3, and 4. This was also explained very carefully before the testing commenced. It is our impression that this change in answer sheets caused little difficulty except for the lowest-scoring students. Some of them seemed to have difficulty with it, but those who did had little likelihood of passing the test (in our estimation). Their difficulty with the answer sheet may, however, have lowered the average score for the group, and that should be taken into consideration in interpreting the results. (The proportion passing was probably not changed at all for this group since their scores were usually far below the cutoff.)

Although we had expected to save time and convenience by using locally-scannable answer sheets on a local facility, we were disappointed in that expectation. We discovered that, for some reason, the sensitivity adjustment

of the local scanner was set at such a level that some marks that were clearly visible to the eye were read as blanks by the local machine. Once that was discovered, all blanks were carefully checked. We found very few errors of that kind, but we made certain that no person was reported as failing a standard due to that kind of error and that no gross failure in scanning could distort the data enough to be of consequence.

As with any testing effort, there are always subjects who become classed as irregular. Usually they are cases who arrive late or become ill during testing. In our testing, the more prevalent irregularities were those who had to leave early before they had completed all they wanted to do. In each of those cases, the examinee upon leaving was asked whether he had finished. If he said that he had, his test was included in the data. If he said that he had not, his paper was not included. In one or two cases, examinees asked to leave apparently to go to the restroom, but did rot return. Their papers were not included in the data. Twelve persons were irregularities of this kind and are not included. The one case from the adult-basic-education group in the pilot study was not included, since he never got as far as the communications part of the Literacy Test. The group who had only the lowest level of the cloze test were not included in analyses involving that

test. As our coders, who transferred the cloze test responses to the machine-readable document, classified the employment level, and similar activities, were working with the answer sheets, they discovered seven cases who had not filled in the demographic sheets or the cloze test. They were eliminated from the study. Finally, when all the data had been transferred to a computer file, it was observed that 22 cases had no responses to either the communications items or the mathematics items, or both. Since these people had not been measured, they were not included in the study. The data of some cases has multiple flaws, so the total tested minus the number deleted for various reasons does not equal the final number of subjects. As a result of this pruning of the cases, a total of 448 cases remained for analysis. Their level and location appear in the chart below, which can be contrasted with the chart earlier presented of all subjects tested.

Description of Subjects

Region	Level of Program	Number	Percent
Northwest	Adult Basic	55	12.3
Northwest	Adult Secondary	73	16.3
North Central	Adult Basic	12	2.7
North Central	Adult Secondary	13	2.9
North Central	Employed Adults	48	10.7
Northeast	Adult Basic	42	9.4
West Central	Adult Basic	53	11.8
Southeast	Adult Basic & Secondary	96	21.4
Southeast	Community College	56	12.5
Totals		448	100.0

Results

Several sets of statistics will help describe the data and the sample. In Tables 2 and 3 appear the means and standard deviations by race (Black, White, and Other) and sex.

Table 4 contains a frequency distribution of the ages of subjects in the total group. Although we asked for adults only to be included at testing centers, in some adult high schools there were students below the age of 18. Apparently these students had stopped going to regular high school and had transferred to adult high school. Thus they were treated as adults.

Table 2

Mathematics and Communications Functional Literacy
Means and Standard Deviations for the Total Group by Race*

Race		Mathematics Mean	Mathematics S.D.	Communications Mean	Communications S.D.
White	N=214	38.832	14.033	46.037	13.774
Black	N=174	22.897	12.266	33.098	15.578
Other	N=47	27.468	14.128	41.872	11.674
Total Group	N=448	31.391	15.502	40.40	15.747

* 13 missing observations due to no record of race

Table 3

Mathematics and Communications Functional Literacy
Means and Standard Deviations for the Total by Sex*

Sex		Mathematics Mean	Mathematics S.D.	Communications Mean	Communications S.D.
Male	N=157	33.134	15.96	40.452	15.461
Female	N=285	30.351	15.074	40.407	15.776
Total Group	N=448	31.391	15.502	40.4	15.747

* 6 missing observations due to no record of sex

Table 4

Frequency Distribution of Ages
of Subjects in Total Group

Age	Frequency	Percent
Under 18	41	9.2
18 - 25	184	41.1
26 - 35	100	22.3
36 - 45	47	10.5
Over 45	65	14.5
No Response	11	2.4
Total	448	100.0

Table 5 contains the frequency distributions of the mathematics and communications scores for the total group. It is clear that the communications scores are more highly skewed than the mathematics scores, with the pileup of scores at the top on the communications test. This will affect the differences between our criterion groups on the communications test--the highest group will not be as far from the middle group as on mathematics because there is not enough top on the communications test to let the best students score as high as they otherwise might. This, of course, is not inappropriate for a test of minimum competencies.

Reliability coefficients were computed for the scores of these subjects using the standard procedure for estimating coefficient alpha, a measure of internal consistency. The alpha coefficient for the mathematics standard was .96 and the coefficient for the communications standard was .98. These coefficients are very high--higher than are usually found in educational tests. They indicate that there is relatively little error in estimating students' scores on these standards. We say relatively because the error is, of course, much greater than the error in making most physical measurements, such as height, length, weight, etc. All educational and psychological measurement has the problem of being much less dependable than physical

Table 5

Frequency Distributions of

Scores for Total Sample

Interval Scores	Mathematics		Communications	
	Frequency	Percent	Frequency	Percent
0-10	47	10.5	30	6.7
11-20	85	19.0	33	7.4
21-30	92	20.5	48	10.7
31-40	80	17.9	75	16.7
41-50	77	17.2	96	21.4
51-59	67	15.0	166	37.1
Total	448	100.1	448	100.0

measurements. But these tests have internal-consistency coefficients as high as or higher than such excellent tests as the College Board's Scholastic Aptitude Test, for example. One caveat should be introduced. Our estimates may be unusually high because our sample is unusually heterogeneous. We deliberately sought people widely differing in socio-economic level, and thereby got people differing widely in their scores on these tests.

Table 6 contains the proportion passing the mathematics standard in our criterion groups. Group A is the group with $5000 or less income and eighth grade or less education. There were 40 in that group. Group B has nine to eleven years of education and income of $5000 to $14,000. There were 75 in that group. Group C has twelve or more years of education and earns $14,000 or more. There were 70 in that group. It can be seen that the proportion passing increases markedly from one group to the next.

Table 7 provides the same information on the same subjects for the communications standard. Again the proportion passing increases markedly from group to group. The chi-square test of significance for each of these tables indicated that the results were statistically significant at the .01 level, so there is very little possibility that results such as these could occur by random error. Thus, this criterion clearly indicates that in this respect, the

Table 6

Proportion Passing Mathematics Standard
In Criterion Groups

Group	Proportion Passing
A	.05
B	.35
C	.77

Table 7

Proportion Passing Communications Standard
In Criterion Groups

Group	Proportion Passing
A	.30
B	.72
C	.87

Functional Literacy Test scores are behaving as such scores should be expected to behave, and these data clearly support the contention that the test has statistical and practical validity in terms of being related to how well people function in our society as it is often measured.

While the proportions passing are of particular interest for the Functional Literacy Test due to the way the standard has been set, it is also of interest to examine the mean total scores (number of items correct) on the tests. For the mathematics standard, those means are presented in Table 8; for communications, in Table 9. Once again, the means increase from one group to another as one would expect. The effect of the skew on the communications score is seen in the fact that the difference between the middle group and the high group on communications is noticeably less than on mathematics. Analysis of variance on these means indicates that they are statistically significantly different from each other at the .01 confidence level and could not reasonably be suspected as having occurred by chance alone.

Since the criterion groups are not the same as the total group, it may be of interest to examine descriptive statistics for criterion groups. Table 10 gives the proportions of Black, White, and Other in each criterion group. Table 11 gives the distribution by ages in each group. Table 12 gives the breakdown by sex in each group.

Table 8

Means and Standard Deviations of
Total Correct Scores on Mathematics in Criterion Groups

	Mean Total Score	
Group	Mean	S.D.
A	18.8	11.3
B	34.3	12.7
C	46.7	13.2

Table 9

Means and Standard Deviations of
Total Correct Scores on Communications in Criterion Groups

	Mean Total Score	
Group	Mean	S.D.
A	27.5	18.0
B	43.6	13.3
C	50.8	11.9

Table 10

Criterion Groups
Description by Race

Groups \ Race	White	Black	Other	No Response	Total
A	6	27	5	2	40
B	38	29	8	0	75
C	62	3	3	2	70
Total	106	59	16	4	185
Percent	57	32	9	2	100

Table 11

Criterion Groups
Description by Age

Groups \ Age	Under 18	18-25	26-35	36-45	Over 45	No Response	Total
A	0	9	10	2	16	3	40
B	13	34	13	8	6	1	75
C	2	24	16	15	12	1	70
Total	15	67	39	25	34	5	185
Percent	8	36	21	14	18	3	100

Table 12

Criterion Groups
Description by Sex

Group \ Sex	Male	Female	Total
A	12	28	40
B	26	49	75
C	32	38	70
Total	70	115	185
Percent	38	62	100

We asked the examinees to tell us what kind of work they did. Our plan was to use these data as part of our definition of criterion groups since often groups are defined for purposes such as these in terms of education, income, and level of occupation. Unfortunately, only 158 of our subjects gave us usable job descriptions. A large number were housewives, and to add them to the data would have required giving them an arbitrary designation of level that probably would only serve to cloud the issues. Housewives can be people of any level of societal functioning. Some of the examinees were essentially full-time students and not working at any job, so there was no good way to classify them either. Preliminary analyses also revealed that the Duncan Decile coefficient was highly redundant with the variables of education and income for the subjects who did report their jobs. It thus could be expected to add little or nothing to analyses, and it was dropped from further study.

The last criterion was the cloze reading test. For this data analysis we used only people who had taken the entire cloze test, and we used all the remaining cases in the test administration, regardless of their age, education, or job level. The total number of cases in the analysis was 413. First, we evaluated the extent to which the cloze levels again provided a Guttman scale. The usual coefficients were calculated with the results appearing in Table 13. All of

Table 13

Scalability Coefficients of Cloze Test Scores
by Literacy Sub-Domains and Combined Sub-Domains
(75% Comprehension)

Sub-Domain	R	MMR	PI	CS	N
Employability Skills	.96	.73	.23	.86	76
Health	.93	.69	.24	.78	87
Money Management	.98	.79	.19	.90	100
Food Preparation	.96	.73	.22	.84	150
Combined Sub-Domain	.96	.73	.23	.85	413

R = Coefficient of Reproducibility
MMR = Minimum Marginal Reproducibility
PI = Percent Improvement (R - MMR)
CS = Percent of Scalability (PI + 1 - MMR)

those values are very high, so the scale characteristics were retained in this set of data, as we had anticipated.

Second, we correlated cloze test scores with the total correct scores on the Functional Literacy Test communications standard for Group B. The correlation was .64, indicating that the communications standard total number right score corresponds very closely to the results obtained from a separate and quite different measure of reading comprehension. (The correlation for the communications standard, pass or fail, was .58.) A validity coefficient as high as this is higher than the level of validity that the College Entrance Examination Board's Scholastic Aptitude Test has for predicting freshman-year grades at most colleges. This adds substantial further support to the concurrent (correlational) validity of the communications standard scores.

Third, and perhaps most interesting, is the analysis which attempts to determine the level of reading represented by the 70% cutting score on the communications standard. To obtain that indication, we used Group B (education grades 9-11, income $5000 to $14,000), and determined the score on the cloze scale that represented the point where the frequency distribution for the failing examinees in that group crossed the frequency distribution for the passing

examinees. This is a standard procedure for making a determination of this kind (Guilford, 1956). It turns out that this point is a scale score on the cloze test of slightly above 3.

People who score at this level have been found to be able to read with at least minimum comprehension material such as the following:

The Confessions of Nat Turner, by W. Styron. Random House, 1967.

Black Like Me, by J. H. Griffin. Houghton Mifflin, 1960

The Hunchback of Notre Dame, by V. Hugo. Dodd, 1947.

Kon Tiki, by T. Heyerdahl. Pocket Books, 1971.

Twenty Thousand Leagues Under the Sea, by J. Verne. Charles Scribner's Sons, 1960.

They will have considerable difficulty with this material and will not usually read it for pleasure, but they can extract information from it. Those who did not pass the communications standard probably cannot extract useful information from material at this level.

Another way to appreciate this level of reading competence is to consider the paragraphs that the students could read with minimal comprehension at this level. One selection, for example, was the following:

The house fly is one of the most common pests found in the home. They breed and feed in garbage, human wastes and animal manure. If disease germs are in these materials, the flies get them on their hairy legs and feet, then fly from these filthy places carrying the germs to your food. House flies have been found to be infected with 20 human diseases including typhoid fever, dysentery and a number of parasitic worms.

To get rid of flies in and around your house, good housekeeping is the most important thing. Keep all garbage in a can with no holes and a tight fitting lid. Wash garbage cans every week with soap and water. Bury, spread, or otherwise dispose of dog, cat, or other animal wastes. Spread lime over wastes in outdoor toilets every few days to reduce fly breeding and odor. Put screens on the windows and doors. Be sure there are no holes in the screens or cracks between them and the frames. Try to make all screen doors open outward. Keep food in covered dishes. Clean all crumbs and food scraps from tables and counters just as soon as you are finished eating.

To kill flies, keep a fly swatter handy.

Most students who did not pass could not make sense of this selection.

Students who just met the passing standard, but with little margin to spare, i.e., who were at level three rather than level four or higher, could be expected to extract useful information only infrequently and with considerable difficulty from the following paragraphs. Those who failed would rarely be able to understand it.

> Babies require very little discipline. They need most your acceptance of them as they are, with tolerance for their immaturity. Their wants are urgent, and they have little ability to postpone their satisfactions. Their needs have to be met fairly soon after becoming known.
>
> After the first few months, your baby may go through the night without a feeding, letting his parents sleep without interruption. Around the first year of life the cup has been gradually introduced, and the bottle discarded. With developing maturity your baby is able to give up the satisfactions of nursing. This is also discipline in part, for he does this because mother wishes to gain freedom from some of the dependency of infancy, and introduce him to substitute feedings. Later, you patiently but

persistently expect him to control his bowels and bladder. This is more discipline; you let the child know he is to take over these controls because you want it and, also, because you wish to recognize his emerging maturity. You maintain your right to eventual freedom from washing the child's soiled clothing, at the same time respecting and encouraging his right to mature--to develop this control--gradually.

Even before he achieves bowel and bladder control, you expect your child to begin to take over his own feeding.

One more kind of information might be helpful in communicating the level of reading represented by the standard on the communications section of the Functional Literacy Test. When the paragraphs were gathered which made up the material from which the cloze test was developed, the selections for level three included, as well as the above paragraph prepared by the Florida Cooperative Extension Service, a paragraph from a pamphlet from the U.S. Department of Agriculture entitled, "Your Money's Worth in Foods," a paragraph from a pamphlet from the New York State Department of Labor entitled, "Why Young People Fail to Get and Hold Jobs," and a paragraph from a brochure from Lewis State Bank in Tallahassee, Florida, entitled, "Savings."

Students at the passing score level on the communications standard of the Florida Functional Literacy Test can extract information from such documents, but will find them too difficult for routine or efficient reading.

Summary

In this study we attempted to investigate whether the scores on the Florida Functional Literacy Test behaved in certain circumstances as one would expect them to behave. This is part of the development of a network of such relationships necessary to establish on a firm foundation the construct validity of the test. Its content validity is well established, and we found its reliability, by the usual standards, is quite high. Over the years findings that the scores behave soundly will support the making of decisions on the basis of this instrument.

We proceeded in this study by locating groups of adults functioning at various levels of proficiency in our society, as judged by their education and income. We gave these people the Literacy Test, asked them to fill out a form describing themselves and their work, and also gave them an independent measure of their ability to read with comprehension.

A substantial correlation was found between the communications total score on the Literacy Test and the independent measure of reading ability. The concurrent

validity is about as high as can ordinarily be expected between two different instruments designed to measure approximately the same thing. This is strong support for the statistical validity of the Literacy Test communications score. We found that a person at the passing level can read such things as the following, not easily but with minimal comprehension.

> To get rid of flies in and around your house, good housekeeping is the most important thing. Keep all garbage in a can with no holes and a tight fitting lid. Wash garbage cans every week with soap and water.

We also found that the proportion passing the mathematics standard and the communications standard increased markedly from the group functioning at a low level in society (income below $5000 and education less than 9th grade) to the group functioning at a high level (income above $14,000 and education of 12th grade and above). The mean scores behaved similarly. This is strong statistical support for the validity of the mathematics and communications scores. They are behaving precisely as one would expect them to if they were measuring a construct that reasonably could be required for high-school graduation when the desire is to grant diplomas to those who have the competencies that will enable them to function effectively in society.

References

1. Guilford, J. P. Fundamental Statistics for Psychology and Education, Third Edition. New York: McGraw-Hill, 1956, pp. 340-342.

2. Guttman, L. The Basis for Scalogram Analysis. In Stouffer, S. A., et al., Studies in Social Psychology in World War II. Vol. IV. Measurement and Prediction. Princeton, N. J.: Princeton University Press, 1950, 60-90.

3. King, F. J. A Content Referenced Interpretive System for Standardized Reading Tests. Final report to the Research Foundation of the National Council of Teachers of English, 1974.

4. Northcutt, N., Selz, N., Shelton, E., and Nyer, L. Adult Functional Competency: A Summary. Unpublished Report. The University of Texas at Austin, 1975.

5. Robinson, John P., Athanasiou, Robert, & Head, Kendra B. Measures of Occupational Attitudes & Occupational Characteristics. Ann Arbor: Survey Research Center Institute for Social Research, University of Michigan, 1969 (Seventh Printing, 1976), pp. 342-356.

* * * APPENDIX * * *

FOR OFFICE USE ONLY	
V: Yes / No	M: Yes / No

1. Name: _____

   Address: _____

   _____

2. What is the highest grade in school you finished?

[ ] 0	[ ] 4th	[ ] 8th	[ ] 12th
[ ] 1st	[ ] 5th	[ ] 9th	[ ] 13th
[ ] 2nd	[ ] 6th	[ ] 10th	[ ] 14th
[ ] 3rd	[ ] 7th	[ ] 11th	[ ] more than 14th

3. Do you have a job?

   [ ] yes
   [ ] no (IF YOU DO NOT HAVE A JOB, GO TO QUESTION 5)

4. What do you do for your job?

   _____

   Is your job: [ ] full time [ ] parttime

5. IF YOU DO NOT WORK, check one of the answers that fits you:

   [ ] I am a housewife
   [ ] I am looking for a job
   [ ] I am not looking for a job

   EVERYONE ANSWER ALL THE REST OF THE QUESTIONS

6. Does anyone else in your house work?

   [ ] yes
   [ ] no

7. What is the total amount of money that you (and everyone else in your house who work) make a year?

   [ ] under $5,000
   [ ] $5,000-$10,000
   [ ] $10,000-$14,000
   [ ] more than $14,000

8. How many children do you have? _____ children

9. Your age:

   [ ] 18-25 years
   [ ] 26-35
   [ ] 36-45
   [ ] over 45

10. Your ethnic group:

    [ ] White
    [ ] Black
    [ ] Hispanic
    [ ] Asian
    [ ] Other _____

11. Sex:

    [ ] Male
    [ ] Female

12. What language do you speak?

    [ ] English
    [ ] Spanish
    [ ] Other _____

VITA

F. J. King

F. J. King is a Professor in the Program of Educational Evaluation and Research Design, Florida State University. His doctoral degree in educational psychology was conferred by the University of Texas (Austin) in 1960. During the year 1957-1960 he was a research associate on the Human Talent Project at the University of Texas; since that time he has been at Florida State.

His teaching responsibilities are in the areas of educational measurement and research design. Current research interests have to do primarily with individual differences and classroom learning. Recent publications include:

With J. K. Brewer. "A Note on a Cognitive Probability Model Incorporating the Serial Position Effect." Psychometrika, 1968, 33(2), 249-252.

With P. Dunn-Rankin. "Multiple Comparisons in a Simplified Method of Scaling." Educational and Psychological Measurement, 1969, 29(2), 315-329.

With D. M. Roberts and R. P. Kropp. "Relationship Between Ability Measures and Achievement Under Four Methods of Teaching Elementary Set Concepts." Journal of Educational Psychology, 1969, 60(3), 244-247.

With P. Dunn-Rankin and Mildred Shimizu. "Reward Preference Patterns in Elementary School Children." International Journal of Educational Science 1969, 3, 53-62.

Aptitude by treatment interactions in mathematical learning. Paper presented in a symposium on aptitude treatment interactios. American Educational Research Association, 1969, Los Angeles, California.

With D. M. Roberts and R. P. Kropp. "An Empirical Investigation of Ferguson's Theory of Human Abilities." Canadian Journal of Psychology, 1969,23(4), 254-269.

With R. M. Vitale, D. W. Shontz, and G. M. Huntley. Effect of Sentence-Combining Exercises Upon Several Restricted Written Composition Tasks. Journal of Educational Psychology. 1971, 62, 521-525.

With Roy O'Donnell. An Exploration of Deep Structure Recovery and Reading Comprehension Skills . Research in the Teaching of English. 1974, 8, 327-339

A Content Referenced Interpretive System for Standardized Reading Tests. Final Report to the Research Foundation of the National Council of Teachers of English, 1974

With D. L. Heinrich, R. S. Stephenson, and C. D. Spielberger. "An Investigation of the Causal Influence of Trait and State Anxiety on Academic Achievement." Journal of Educational Psychology. 1976, 68, 330-334.